

Małgorzata Janiak

## **GRID – DYNAMICZNY DOSTĘP DO ZASOBÓW**

### **[GRID – DYNAMIC NETWORK]**

**Abstrakt:** Artykuł przedstawia zjawisko *Grid*, czyli siatki złożonej zarówno z zasobów bazodanowych, jak i sprzętowych. System ten zapewnia naukowcom dostęp do usług zdalnego udostępniania kosztownych urządzeń laboratoryjnych lub obliczeniowych oraz do usług zdalnej wizualizacji danych. Umożliwia także dostęp do wielu źródeł informacji o aktualnie prowadzonych badaniach lub o opracowanych już wynikach.

GRID – SIEĆ DYNAMICZNA

**Abstract:** This paper presents a new constructivist model of knowledge development – *Grid*. The *Grid* is widely seen as the next generation Internet. This system aims to share dynamic collections of individuals, institutions and resources by providing consistent, easy and inexpensive access to high-end computational capabilities.

DYNAMIC NETWORK – GRID

\*

\*      \*

„Nadmiar informacji” wyprodukowanej przez ludzi to określenie, które stało się dla niektórych wyświechtanym i nadużywanym frazesem. W celu zapewnienia użytkownikom dostępu do tak dużej ilości informacji powstał m.in. Internet. Nie spełnił jednak wszystkich oczekiwań, zwłaszcza tych, jakie pokładali w nim naukowcy.

Badacze testują i pracują więc nad nowym modelem sieci, który nie ma być kontynuatorem Internetu bądź tzw. *Open Systems*, ale całkiem nowym systemem zapewniającym

---

Mgr Małgorzata JANIĄK, wykładowca w Instytucie Informacji Naukowej i Bibliotekoznawstwa UJ. Adres elektr.: [malgosia@inib.uj.edu.pl](mailto:malgosia@inib.uj.edu.pl)

dostęp do olbrzymiej ilości już zgromadzonych oraz ciągle tworzonych, nowych zasobów. Dostęp do tej sieci, zwanej *Grid*, możliwy jest przez Internet, ale także przez telefon komórkowy lub skomplikowane komputerowe wizualizacje.

*Grid*, czyli tzw. Siatka, to (opisowo i quasi-definicyjnie) infrastruktura umożliwiająca współużytkowanie zasobów informacyjnych oraz komputerowych, które pozostają rozproszone geograficznie. To rozproszenie i podległość różnym domenom administracyjnym jest jedną z podstawowych cech tego systemu. Drugą ważną cechą jest brak scentralizowania źródeł, które są jedynie koordynowane, ale otwarte na ustandaryzowane protokoły komunikacyjne. Architektura oprogramowania *Grid*-owego została już zdefiniowana w standardzie OGSA (*Open Grid Services Architecture*). Innym opracowanym standardem jest WSRF (*Web Services Resource Framework*), który określa otoczenie i sposób budowania takiego oprogramowania z wykorzystaniem usług sieciowych (*Web Services*) [Grid 2005, dok. elektr.]. Trzeci wyróżnik *Grid*, według Iana Fostera (o którym za chwilę), to fakt, iż dostarczane usługi mają „różnicowany i nietrywialny charakter”, czyli nie mogą wymagać od użytkownika specjalistycznej wiedzy i skomplikowanych procedur [Dziwisz 2005, dok. elektr.]. Całe zasoby Siatki są zaś dla użytkownika przezroczyste, tzn. użytkownik nie musi wiedzieć, skąd pochodzą udostępniane mu dane, na jakim serwerze dokonywane są obliczenia, ani które dokładnie fragmenty infrastruktury są angażowane w wykonanie jego zadania.

Sam termin *Grid* odnosi się do przetwarzania siatkowego (*Grid computing*), którego zaletą są różne rozwiązania klastrowe, szybsze sieci i współużytkowana pamięć masowa. Samo przetwarzanie siatkowe zaś oparte jest na modelach oddzielnych grup użytkowników, którzy dynamicznie użytkują różne zasoby informatyczne poprzez sieci o dużej przepustowości, umożliwiające elastyczne dostosowanie do zmieniających się potrzeb [Miley 2005, dok. elektr.].

Michael Cosnard przełożył angielski *Grid* na język francuski, tworząc z terminu akronim od: *Globalisation de ressources informatiques et données* [Kusznik 2005, dok. elektr.], co trafnie oddaje zakres znaczeniowy.

Próba zdefiniowania terminu w Wikipedii przedstawia się następująco:

Siatka komputerowa (często także z ang. Grid) to system, który integruje i zarządza zasobami będącymi pod kontrolą różnych domen (od instytucji po system operacyjny) i połączonymi siecią komputerową, używa standardowych, otwartych protokołów i interfejsów ogólnego przeznaczenia (odkrywania i dostępu do zasobów, autoryzacji, uwierzytelniania) oraz dostarcza usług odpowiedniej jakości (QoS, oferuje usługi wyższego poziomu). (...) Siatka jest rozwinięciem idei klastra poza tradycyjne granice domeny [Grid 2005, dok. elektr.].

Początki *Grid* łączone są z Ianem Fosterem [zob. <http://www-fp.mcs.anl.gov/~foster/>], obecnym dyrektorem Distributed Systems Lab Mathematics & Computer Science Argonne National Laboratory (Argonne, IL – USA) oraz z Carlem Kesselmanem [zob. <http://www.isi.edu/~carl/>] z USC/Information Sciences Institute (Marina del Way, CA – USA), obecnie dyrektorem Center for Grid Technologies tego Instytutu.

Historycznie rzecz biorąc, początki Siatek to wczesne lata 90., kiedy to w USA, dzięki gigabitowym instalacjom testowym CASA i BLANCA, połączono wiele rozproszonych po

kraju laboratoriów. Uznaje się, że pierwsza próba przetwarzania danych w sieci typu *Grid* to tygodniowy eksperyment, przeprowadzony w 1995 r., w czasie którego naukowcy ze Stanów Zjednoczonych połączyli komputery znajdujące się w siedemnastu różnych miejscach (w ramach projektu *I-Way*).

Celem *Grid* jest przede wszystkim implementacja nowoczesnych, szerokopasmowych usług zdalnego udostępniania zasobów komputerowych, ale także samych kosztownych urządzeń laboratoryjnych, z wykorzystaniem znanych już interfejsów lub protokołów. Badacze oczywiście pracują też nad ulepszeniem i tworzeniem nowych mediów.

Implementacja ta daje możliwości przeprowadzania obliczeń w dużej skali oraz tworzenia bardzo zaawansowanych wizualizacji. Dzięki takiej siatce każdy użytkownik z tzw. Wirtualnych Organizacji (o których będzie mowa później) ma przezroczysty dostęp do udostępnianych zasobów, w tym także do istniejących obecnie i rozbudowywanych w przyszłości różnorodnych systemów przechowywania danych.

Środowiska *Grid* tworzą przede wszystkim trzy główne typy składników:

- stacje robocze (*scavenging Grids*), czyli komputery osobiste, serwery instytucji, wyposażenia laboratoryjne itp., które nie są wykorzystywane stale w całej ich mocy obliczeniowej, w związku z czym mogą być udostępnione innym użytkownikom;
- serwery (*computational Grids*), czyli połączone komputery o dużej mocy obliczeniowej, które umożliwiają efektywną realizację bardzo złożonych zadań przetwarzania danych;
- dane (*data Grids*), których główną cechą stało się współdzielenie ich w trakcie realizacji przetwarzania *Grid Computing*; poszczególne aplikacje korzystają z baz danych lub wirtualnego systemu plików, których lokalizacje są przezroczyste dla aplikacji [Zakrzewicz 2005, dok. elektr.].

Rynek oprogramowania oferuje bardzo dużą liczbę narzędzi wspomagających siatki. Najpopularniejszymi z nich są: Globus Toolkit (Globus), Oracle Globus Development Kit (oparty na Globus-Oracle), IBM Grid Toolbox (oparty na Globus-IBM), Avaki (Avaki), DataSynapse LiveCluster (Synapse), Entropia DCGrid (Entropia), Platform LSF, ActiveCluster oraz MultiCluster (Platform), United Devices MetaProcessor Platform (United) [Zakrzewicz 2005, dok. elektr.].

Założenia siatki to, według Benny'ego Soudera z firmy Oracle:

wdrożenie systemów, które umożliwią dynamiczne przydzielanie zasobów, elastyczne dostarczanie informacji i dużą dostępność. (...) Dynamiczne przydzielanie zasobów oznacza gwarancję, że zasoby nie pozostaną bezczynne, gdy żądania użytkownika czekają na obsługę, ale będą szybko przydzielane na żądanie. Elastyczne dostarczanie informacji równa się pewności, że jest ona dostępna użytkownikom i aplikacjom tam, gdzie tego potrzeba i w odpowiednim czasie. Przez dużą dostępność rozumiemy, że dane i możliwości obliczeniowe zawsze muszą być dostępne, tak jak prąd elektryczny [Miley 2005, dok. elektr.].

Dostęp do sieci zapewniają użytkownikowi Wirtualne Organizacje (*Virtual Organizations*), tworzone przez różnego rodzaju badaczy oraz przez instytucje zainteresowane współpracą lub połączone wspólnymi celami. Obecnie najbardziej zainteresowani są przedstawiciele nauk ścisłych. *Grid* wykorzystywany jest już dla potrzeb bioinformatyki, astro-

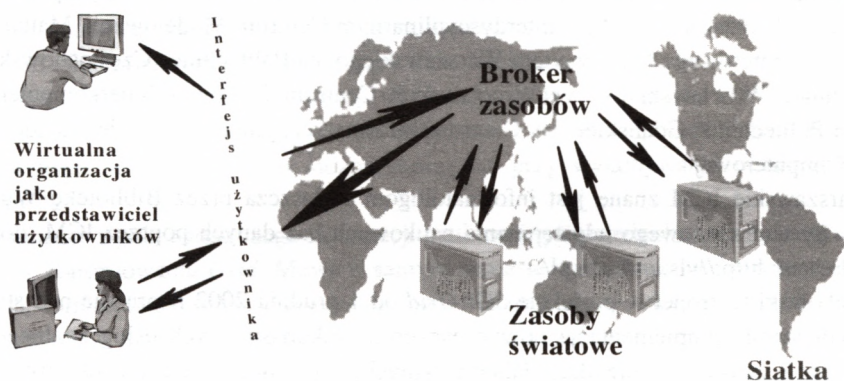
nomii, obserwacji ziemi, badań fizycznych, chemicznych, medycznych oraz w meteorologii. W dziedzinach tych ilość danych bywa gigantyczna. Przykładowo przy obserwacji ziemi naukowcy otrzymują i powinni przetworzyć 100 GB danych dziennie. Dzięki *Grid* mogą obróbkę informacji prowadzić poprzez komputery na całym świecie. Mają też możliwość porównywania wyników z informacjami, zgromadzonymi na całym świecie, o ile tylko komputery podłączone są przez jakiś *Grid middleware* (software przygotowany dla potrzeb siatki).

*Grid* służy jednocześnie także przy *data minig*, czyli bardzo zaawansowanym wyszukiwaniu informacji w wielu dużych bankach danych, często rozproszonych lub niespójnych pod względem struktur, formatów lub języków. Siatka pomaga też w indeksowaniu baz medycznych, które są doskonałym przykładem baz skomplikowanych strukturalnie, dużych ze względu na ilość danych oraz często wykorzystywanych wielojęzkowo przez użytkowników z różnych krajów.

Każda nowo powstała Wirtualna Organizacja jest sprawdzana przed jej włączeniem w struktury Siatki, co ma zapewnić ochronę posiadanych danych oraz komputerów przed np. różnego rodzaju wirusami. Globalne Forum Gridowe (*Global Grid Forum*), gdzie użytkownicy mogą znaleźć wiele potrzebnych informacji, znajduje się pod adresem: <http://www.Gridforum.org>.

Użytkownik łączy się poprzez organizację z tzw. brokerem zasobów (*Resource Broker*), który przydziela odpowiednie komputery do wykonania obliczeń bądź bazy danych do przeprowadzenia wyszukiwań. Serwis ten szuka wolnych zasobów oraz ustawia kolejkę dla konkretnego pytania lub zadania obliczeniowego.

Rysunek 1. Struktura Siatki *Grid*



Światowi użytkownicy komputerów i sieci po raz pierwszy wykorzystali taką Siatkę w latach 2003–2005 w programie SETI@Home, przypominającym *Grid*. Miliony osób ściągnęło program, który przetwarzał i wysyłał dane z radioteleskopów poszukujących istot inteligentnych poza Układem Słonecznym. Program ten nie obciążał komputera konkretnie-



go użytkownika (tzw. stacji roboczej) w trakcie jego pracy, gdyż włączał się, gdy osoba przestawała pracować na swoim komputerze [zob. [http://www.setiathome.pl/home\\_polish.shtml](http://www.setiathome.pl/home_polish.shtml)].

Obecne inicjatywy, tzw. *World Community Grid*, dotyczą m.in.: badań nad AIDS, nowotworami, chorobą Alzheimera, przewidywania kataklizmów, ochrony światowych zasobów wody i żywności. Pochodzą one od firmy IBM, ONZ, badaczy z Mayo Clinic, Oxford University i innych. Informacje na ten temat znaleźć można w Internecie pod adresem: <http://www.worldcommunityGrid.org> [*Społeczny...* 2005, dok. elektr.].

Laboratorium CERN także wykorzystuje już cudze zasoby: procesory, laboratoria i bazy. Informacje na ten temat znaleźć można pod adresem: siatka LHC <http://otn.oracle.com/oramag/oracle/cernGrid>. Inne zaś opisane zostały na liście kluczowych projektów związanych z siatkami (według Iana Fostera) pod adresem: <http://www.mcs.anl.gov/~foster/Grid-projects>. W Europie główną rolę odgrywa sieć *GÉANT* łącząca 26 sieci badawczych. Informacje na jej temat znaleźć można pod adresem: <http://www.dante.net/server/show/nav.007>.

Tworzone są także coraz to nowe projekty badawcze. Komisja Europejska we współpracy z USA i Rosją zaakceptowała projekt EGEE – *Enabling Grids for E-Science in Europe* w 6. Programie Ramowym (grantowym). Przyjęto do niego 71 dotowanych przez Komisję Europejską partnerów oraz około 30 dodatkowych instytucji współpracujących. Głównym koordynatorem zadań jest CERN. Trzy zasadnicze pola działań to: *NA – Networking Activities* (nowe rozwiązania dla sieci), *SA – Specific Service Activities* (działania nad tworzeniem i zarządzaniem serwisami) oraz *RA – Research Activities* (metody badań) [Kusznik 2005].

W Polsce badania nad wykorzystaniem *Grid* prowadzone są: w Poznańskim Centrum Superkomputerowo-Sieciowym (PCSS), w ACK (Akademickim Centrum Komputerowym) Cyfronet w Krakowie, w ICM – Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, na Politechnice Częstochowskiej, na Politechnice Wrocławskiej – we Wrocławskim Centrum Sieciowo-Superkomputerowym oraz na Politechnice Gdańskiej – w Centrum Informatycznym Trójmiejskiej Akademickiej Sieci Komputerowej.

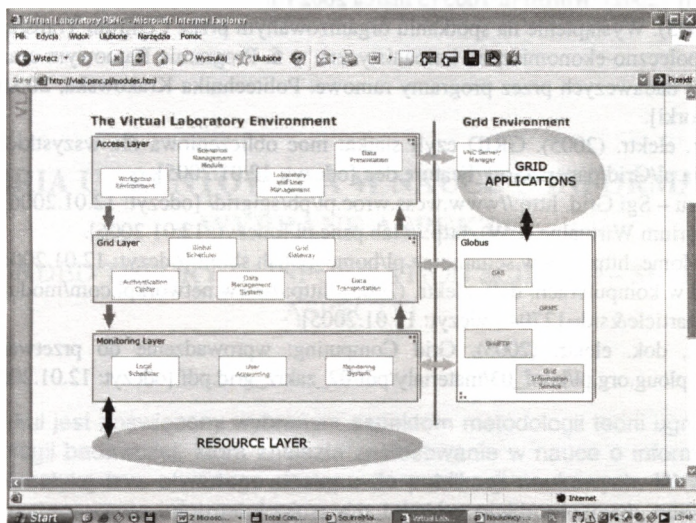
Warszawskie ICM znane jest informatologom zwłaszcza przez Bibliotekę Wirtualną ICM – System sieciowego udostępniania naukowych baz danych poprzez ICM, dostępną pod adresem: <http://vls.icm.edu.pl/>.

Krakowski Cyfronet w projekcie *SGI Grid* od 1 grudnia 2002 r. pracuje nad stworzeniem koncepcji i implementacją nowoczesnych szerokopasmowych usług zdalnego udostępniania kosztownych urządzeń laboratoryjnych, zapasowego centrum obliczeniowego oraz usługi zdalnej wizualizacji danych. Opis projektu znaleźć można pod adresem: <http://www.wcss.wroc.pl/pb/sgigrd>.

Pod koniec 2005 r. (także w grudniu) Poznańskie Centrum Superkomputerowo-Sieciowe przygotowało Laboratorium Wirtualne (LW – <http://vlab.psnc.pl/>), które pozwala naukowcom na korzystanie z przyrządów laboratoryjnych poprzez Internet. Wśród nich zna-

laży się m.in. spektrometr NMR, radioteleskop oraz mikroskop elektronowy. Poza tym Laboratorium Wirtualne pozwala na połączenie tych urządzeń z serwerami obliczeniowymi. Daje to możliwość tworzenia pełnych scenariuszy pomiarowych, umożliwiających przeprowadzenie całego badania w sposób optymalny pod względem czasu i wykorzystanych zasobów. Wyniki zaś mogą być umieszczane w Naukowej Bibliotece Cyfrowej (NBC), która poza przechowywaniem danych pomiarowych umożliwia także ich przeszukiwanie, aktualizację, oraz przede wszystkim współdzielenie się nimi z innymi badaczami.

Rysunek 2. Struktura Laboratorium Wirtualnego



*Grid* jest na pewno zjawiskiem, którym powinna zainteresować się bliżej nauka o informacji naukowej. Na razie widać duże zainteresowanie ze strony informatyki i nauk ścisłych. Ale Siatka jako system obejmujący dane istniejące i ciągle nowo tworzone jest niesłychanie bogatym źródłem informacji, dotychczas tylko dla badaczy z zakresu nauk ścisłych, lecz zapewne niedługo także dla przedstawicieli nauk społecznych. Z możliwości oferowanych przez ten system można skorzystać przy wykonywaniu obliczeń informetrycznych.

Elementy z pola badawczego informacji naukowej już niejednokrotnie były wykorzystywane w konstruowaniu *Grid*. Mowa o samym procesie wyszukiwania danych, struktur baz danych, formatów danych, indeksowaniu, budowie interfejsów wielojęzycznych itp. Rozległość zasobów Siatki oraz jej różnorodność doczekała się także opracowań z perspektywy metadanych – zob. np. referat *Grid Database i Meta informacja* opracowany przez Wydział Matematyki i Nauk Informacyjnych Politechniki Warszawskiej, dostępny pod adresem: [http://www.si.pjwstk.edu.pl/seminaria/prezentacje/Grid\\_architektura\\_a\\_metadane.pdf](http://www.si.pjwstk.edu.pl/seminaria/prezentacje/Grid_architektura_a_metadane.pdf).

Tak więc *Grid* jest znaczącym ułatwieniem w dostępie do wiedzy naukowej zgromadzonej w zasobach ogólnościatowych i jako taki powinien znaleźć się w obrębie zainteresowań nauki o informacji naukowej.

## Wykorzystane źródła i opracowania

- Dziwisz, J., dok. elektr. (2005). Enter the Grid! Noc Linuxożerców. <http://www.noc.linux.pl/nl3/Grid.pdf> [odczyt: 12.01.2005].
- Foster, I.; C. Kesselman; J.N. Nick; S. Tuecke (2002). Grid Services for Distributed System Integration. *Computer* Vol. 35 (6), p. 37–46.
- Foster, I.; C. Kesselman; S. Tuecke, dok. elektr. (2001). The Anatomy of the *Grid*: Enabling Scalable Virtual Organizations. <http://citeseer.ist.psu.edu/cache/papers/cs/25047/http:zSzzSzwww.mcs.anl.govzSzglubuszSzresearchzSzpaperszSz anatomy.pdf/foster01anatomy.pdf> [odczyt: 12.01.2005].
- Grid, dok. elektr. (2005). [http://pl.wikipedia.org/wiki/Grid\\_komputerowy](http://pl.wikipedia.org/wiki/Grid_komputerowy) [odczyt: 12.01.2005].
- Komputer totalny (2002). *Wprost* nr 1005 (3 marca 2002 r.).
- Kusznic, A. (2005). Wystąpienie na spotkaniu organizowanym przez Centrum Transferu Technologii nt. Nauki społeczno-ekonomiczne i humanistyczne w 6. Programie Ramowym oraz finansowanie infrastruktur badawczych przez programy ramowe. Politechnika Krakowska, 20 stycznia 2005 r. [notatki autorki].
- Miley, M., dok. elektr. (2005). GRID czyli siatka: moc obliczeniowa dla wszystkich. [http://www.visual-media.pl/Grid/papers/Grid\\_Feature.doc](http://www.visual-media.pl/Grid/papers/Grid_Feature.doc) [odczyt: 12.01.2005].
- Serwis Cyfronetu – Sgi Grid. <http://www.wcss.wroc.pl/pb/sgigrid/> [odczyt: 12.01.2006].
- Serwis Laboratorium Wirtualne (LW). <http://vlab.psnc.pl> [odczyt: 12.01.2006].
- Serwis SETI@Home. [http://www.setiathome.pl/home\\_polish.shtml](http://www.setiathome.pl/home_polish.shtml) [odczyt: 12.01.2006].
- Spółeczny *Grid* w komputerach, dok. elektr. (2005). <http://www.networkpl.com/modules.php?name=News&file=article&sid=12704> [odczyt: 12.01.2005].
- Zakrzewicz, M., dok. elektr. (2005). Grid Computing: wprowadzenie do przetwarzania danych. [http://www.ploug.org.pl/konf\\_03/materialy/pdf/02\\_zakrz\\_grid.pdf](http://www.ploug.org.pl/konf_03/materialy/pdf/02_zakrz_grid.pdf) [odczyt: 12.01.2005].